



**FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO SEMARANG**

Jl. IMAM BONJOL NO. 207 SEMARANG Telp. 024-3575915, 024-3575916

JAWABAN UJIAN TENGAH SEMESTER GANJIL 2022/2023

Mata Kuliah : Analitika Media Sosial	Sifat : Take Home
Hari/tanggal : Jumat, 3 November 2023	Waktu : 10.20 – 12.00
Kelompok : A12.6503	Dosen : Ika Novita Dewi, MCS
NIM : A12.2021.06620	Nama : Chusnuut Tacharri

A. Judul

Analisis Sentimen komentar Youtube terhadap Pemberhentian Reklamasi Pantai Menggunakan Algoritma Naïve Bayes

B. Pendahuluan

Reklamasi pantai merupakan suatu proses untuk menambah daratan ke arah laut dengan cara menimbun laut dengan material tertentu seperti batu, pasir, dan tanah dengan maksud dan tujuan tertentu yang berkaitan dengan keperluan yang dibutuhkan oleh penduduk. Proyek reklamasi pantai sudah dilakukan sejak zaman era Soeharto di Teluk Jakarta dan telah ditetapkan pada Keputusan Presiden Nomor 52 Tahun 1995 tentang Reklamasi Pantai Utara Jakarta dan Perda Nomor 8 Tahun 1995. Keputusan adanya proyek reklamasi pantai di Jakarta ditentang oleh Kementerian Lingkungan Hidup karena memiliki dampak yang besar. Namun pada pemerintahan Presiden Susilo Bambang Yudhoyono (SBY) hingga pemerintahan Gubernur Basuki Tjahaja Purnama (Ahok) tetap melanjutkan proyek reklamasi pantai tersebut.

Pada era Kepemimpinan Gubernur DKI Jakarta Anies Baswedan, izin pembangunan proyek reklamasi pantai di teluk Jakarta dihentikan pada tanggal 26 September 2018. Dampak dari pemberhentian proyek tersebut yaitu banyak bangunan yang mangkrak dan tidak berpenghuni. Hal itu juga mengundang komentar netizen pada kanal Youtube dari berbagai akun atas pemberitaan diberhentikannya proyek pembangunan reklamasi pantai di Teluk Jakarta. Dengan demikian, ada 4 pulau yang dilanjutkan pembangunannya karena sudah terlanjur yaitu pada Pulau C dan D (pemegang izin PT Kapuk Naga Indah), Pulau G (pemegang izin PT Muara Wisesa Samudra), dan Pulau N (pemegang izin PT Pelindo II).



**FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO SEMARANG**

Jl. IMAM BONJOL NO. 207 SEMARANG Telp. 024-3575915, 024-3575916

Berdasarkan pemberitaan pemberhentian proyek reklamasi pantai tersebut melalui kanal Youtube, dapat mengundang opini netizen dari berbagai akun untuk memberikan komentar mengenai masalah tersebut. Dari komentar netizen itu lah dapat dianalisis dengan cara analisis sentiment yang bertujuan untuk memperoleh opini baik negative atau positif dari adanya pemberhentian proyek reklamasi pantai tersebut serta analisis word cloud untuk memperoleh kata-kata apa saja yang seiring muncul saat digunakan netizen ketika berkomentar berdasarkan masalah tersebut. Dari hasil analisis tersebut dapat diketahui bagaimana opini netizen tentang pemberhentian proyek reklamasi pantai di Teluk Jakarta.

C. Metode Penelitian

1. Pengumpulan Data

Pengumpulan data untuk melakukan analisis sentiment dan word cloud yaitu pada komentar sosial media Youtube dengan link sebagai berikut <https://www.youtube.com/watch?v=q8FDOGWFguE&list=WL&index=2&t=87> serta pengumpulan data juga diambil dari Netlytic dan menghasilkan berupa file dataset.

2. Text Preprocessing

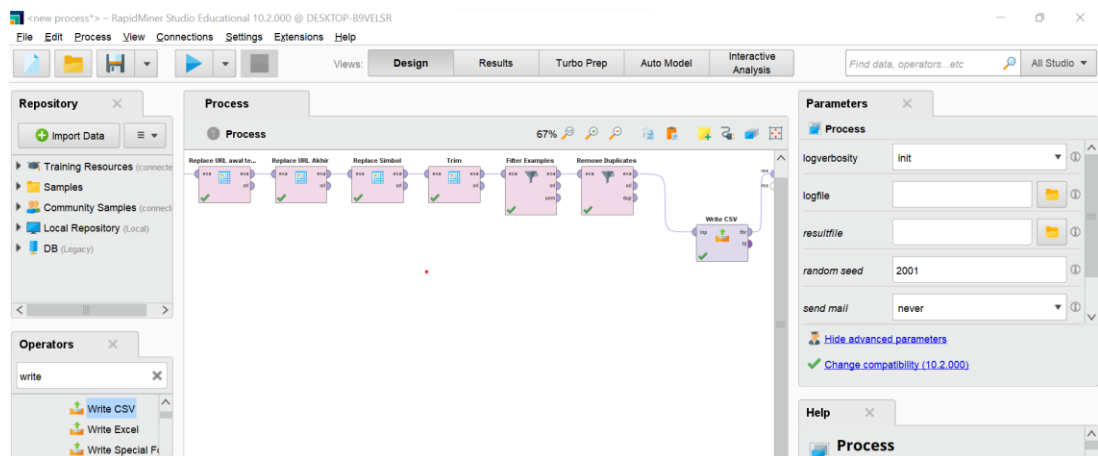
Langkah-langkah dalam membersihkan data text sebagai berikut:

- a. Download dataset pada netlytic.org dalam bentuk excel.
- b. Buka file dataset tadi kemudian pisahkan kolom deskripsi saja dengan membuka file excel baru. Memisahkan deskripsi yaitu untuk mengolah data komentar dari netizen.
- c. Pada atribut description diubah menjadi attribute "Text" kemudian simpan file tersebut dengan format .csv
- d. Kemudian, buka RapidMiner pada Operator cari menu Read CSV, kemudian drag menu tersebut ke bagian process. Setelah itu klik 2 kali pada kotak read CSV kemudian masukkan file CSV tadi yang telah disimpan sesuai dengan file pada poin c.
- e. Setelah itu menggunakan operator replace berfungsi untuk mengubah atau menghapus karakter atau kata tertentu yang akan diinputkan. Operator replace ini dapat langsung dicari pada menu operator dengan mengetikkan "Replace".
- f. Pada text preprocessing, pertama akan menghapus karakter mention @ dengan memasukkan pada kolom replace what sebelah kanan yaitu dengan menginputkan

“@.* “Fungsi .*[spasi] yaitu akan menghapus semua karakter mention @ pada pemrosesan text tersebut. Untuk mengecek apakah sudah berhasil dihapus yaitu dengan menarik read CSV pada out ke Replace exa dan replace exa ke res. Kemudian Run. Setelah itu cek pada tombol result apakah @ sudah berhasil dihapus atau belum. Jika sudah tidak ada maka process replace pada mention @ telah berhasil.

- g. Kedua akan menghapus URL dengan menggunakan operator replace. Pada replace URL ini dilakukan 2x karena meraplace URL yang letaknya di awal, tengah dan url yang letaknya di belakang. Cara mereplace url di awal dan tengah kalimat yaitu dengan menginputkan pada replace what “http.*?[spasi]. Sedangkan untuk mereplace http di belakang kalimat dengan cara menginputkan https.*?. Kemudian ditarik exa-exa dan exa res. Namun jika ingin membandingkan dengan data yang asli maka ditarik replace url ori-res.
- h. Ketiga akan menghapus simbol seperti emoticon dan lain-lain yang tidak penting, yaitu masih menggunakan operator replace dengan menginputkan replace what [!~?.,;:'"#*%@%\$&()-]. Setelah itu tarik pada exa-exa dan exa-res dan jika ingin membandingkan file sebelumnya maka dapat ditarik juga ori-res.
- i. Selanjutnya menggunakan operator trim yang berfungsi untuk menghapus white space atau sepasi-spasi yang tidak perlu. Operator trim dapat langsung dirun dengan menarik exa-exa dan exa-res. Untuk membandingkan dengan file sebelumnya dapat menarik juga ori-exa.
- j. Selanjutnya menggunakan operator filter example yaitu untuk menghapus data kosong atau data yang dapat ditandai dengan “?” tanda tanya saja. Pada filter example dengan menambah filter (add filter) dan memilih is not missing karena akan mengambil data yang tidak kosong. Kemudian dijalankan atau Run.
- k. Langkah terakhir pada text processing ini yaitu menggunakan operator remove duplicat. Pada operator ini berfungsi untuk menghapus data ganda atau data yang sama namun akan diambil satu data saja dari data yang sama dengan memilih atribut filter single dan atribut values “Text” sehingga dari data ganda tersebut akan hanya diambil satu saja.

- l. Setelah data berhasil dibersihkan, selanjutnya simpan data tersebut pada file CSV dengan mencari operator write CSV kemudian ketuk 2x operator tersebut lalu simpan pada folder yang diinginkan. Jika berhasil disimpan maka text preprocessing telah selesai.
- m. Ini merupakan lampiran langkah-langkah dari data Text Preprocessing



3. Document Processing

Langkah-langkah pemrosesan data text:

- a. Masih menggunakan dari proses data preprocessing sebelumnya, selanjutnya memilih operator nominal to text yaitu mengubah data berupa nominal yang akan dikonversikan ke data berbentuk text agar data tersebut dapat diproses. Pada operator nominal to text ini prosesnya dihubungkan dari operator remove duplicate.
- b. Selanjutnya operator process document from data yaitu operator khusus yang memiliki sub-process didalamnya jika diklik 2x maka terdapat operator-operator pemrosesan text.

Berikut ini merupakan operator-operator dari sub-process:

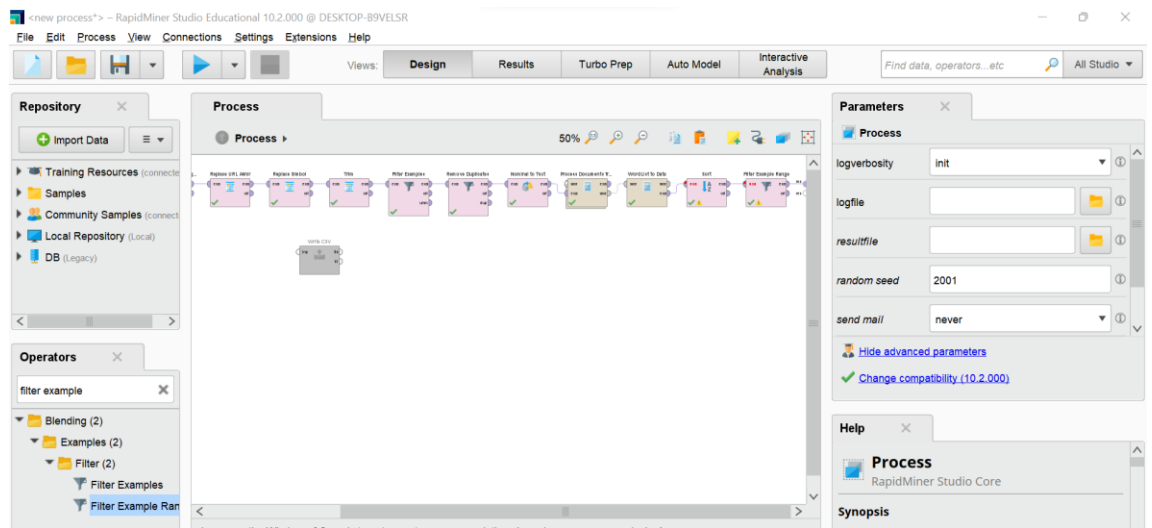
- Operator Tokenize berfungsi untuk memisahkan satu kata dengan kata lainnya sehingga setiap kata akan dipisahkan menjadi satu atribut tersendiri.
- Operator Transform Case berfungsi untuk mengubah font agar semua data bisa diseragamkan menjadi satu missal dengan lower case maka data yang muncul menggunakan huruf kecil semua.



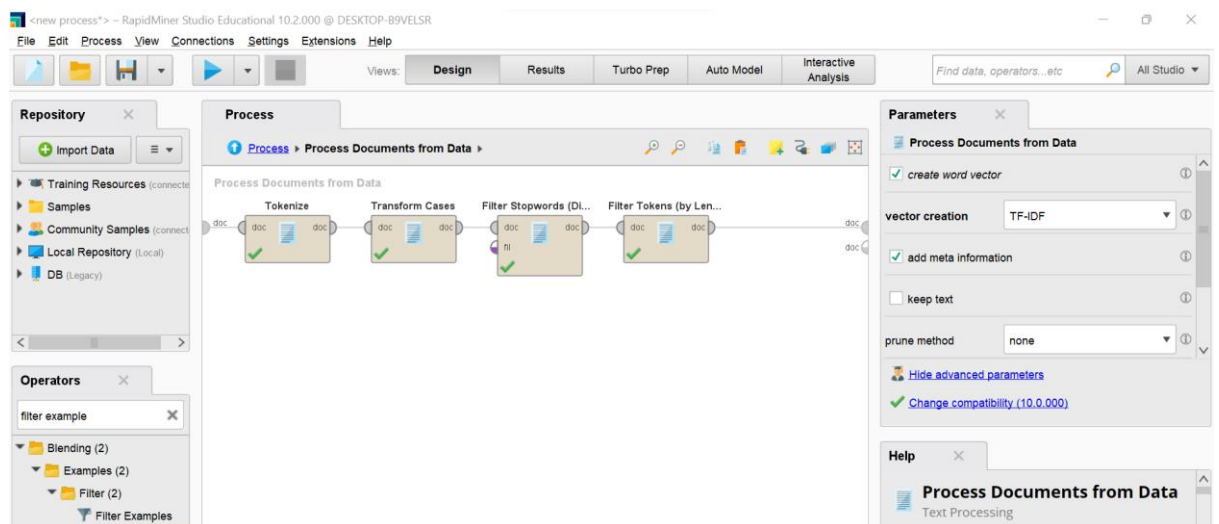
**FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO SEMARANG**

Jl. IMAM BONJOL NO. 207 SEMARANG TELP. 024-3575915, 024-3575916

- Filter Stopwords (Dictionary) karena pada penggunaan kata ini menggunakan bahasa Indonesia oleh karena itu menggunakan dictionary. Filter stopwords ini berfungsi untuk menghapus kata-kata yang tidak penting seperti yang, dan dari, di, adalah dan lain-lain. Namun pada filter ini tidak perlu membuat list sendiri tapi dapat mendownload file stopwords bahasa di kagle.
 - Filter Tokens by Length berfungsi untuk menghapus kata-kata yang mengandung sedikit huruf dan kata-kata yang mengandung banyak huruf. Dan hal itu bisa diatur dengan min 4 maksudnya menghilangkan kata-kata mengandung 4 huruf dan max 25 yaitu menghilangkan kata-kata yang mengandung 25 huruf.
- c. Selanjutnya menggunakan operator WordList to Data untuk menampilkan visualisasi data. Namun output process document data berupa wordlist bukan example sehingga harus dikonversi terlebih dahulu output tersebut menjadi example. Hal itu untuk menampilkan visualisasi data.
- d. Selanjutnya menggunakan operator sort untuk mensortir kata yang dibutuhkan saja atau kata yang sering muncul pada kolom komentar yaitu dengan mensortir descending yaitu mengurutkan kata dari yang terbesar hingga terkecil.
- e. Ketika ingin memvisualisasikan tidak bisa karena data terlalu banyak sehingga membutuhkan operator filter example range dengan memilih sekitar 30 kata yang teratas yang nantinya kata tersebut akan divisualisasikan sehingga dapat mengetahui wordcloudnya atau kata yang sering muncul pada kolom komentar. Dari wordcloud tersebut dapat dijadikan kesimpulan mengenai permasalahan diberhentikannya proyek reklamasi pantai.
- f. Pada visualisasi kata diambil diambil dari kolom ke-3 karena jika kolom 1 diambil maka sudah pasti kata terbanyak karena kata tersebut merupakan keyword dari video youtube. Dan memvisualisasikan menggunakan weight total untuk mengetahui kata apa yang sering muncul
- g. Lampiran langkah-langkah document processing (WordCloud)



h. Lampiran Sub-process dari Process document



4. Text labelling

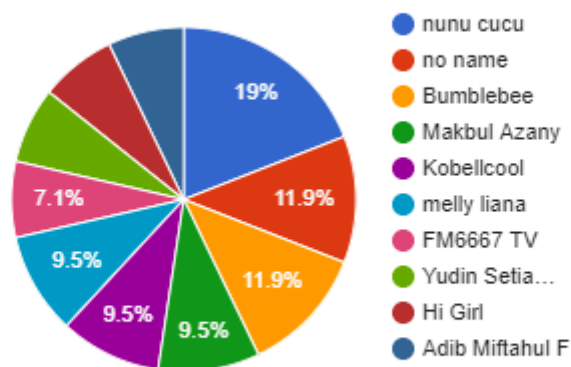
Pada bagian proses text labelling yaitu lanjutan dari proses pada poin C2 dimana hasil data yang sudah dibersihkan akan diberi label yang bernilai positif dan negatif sebanyak 300 baris. Tujuan labelling untuk menganalisis sentiment pada kolom komentar Youtube dan sebagai bahan untuk melakukan proses sentiment analisis dilangkah selanjutnya (sisa dari baris yang belum diberi label).

D. Hasil dan Pembahasan

1. Analisis Top User

Berdasarkan hasil data dari netlythic.org, terdapat 10 akun yang yang paling banyak berkoemntar dalam video tentang nasib proyek reklamasi pantai di Teluk

Jakarta. Berikut ini 10 aktor yang menjadi top users dalam berkomentar pada video proyek reklamasi pantai, antara lain nunu cucumu (19%), no name (11,9%), Bumblebee (11,9%), Makbul Azany (9,5%), Kobellcool (9,5%), melly liana (9,5%), FM6667 TV (7,1%), Yudin Setiawan (7,1%), Hi Giri (7,1%), Adib Miftahul F (7,1%).



Gambar Top User

Pada user nunu cucu merupakan top user karena pada data tersebut menghasilkan 19% dimana user nunu cucu telah mereply pada komentar aktor lain sebanyak 8 kali dan akun ini disebut oleh aktor lain sebanyak 1 kali serta akun nunu cucu meninggalkan komentar 1 kali. Akun nunu cucu mereply komentar orang lain karena akun nunu cucu setuju dengan pemberitaan pemberhentian proyek reklamasi pantai dan masih ada kaitannya dengan perizinan. Dan karena akun nunu cucu dalam komentar selalu menyuarakan tentang perizinan reklamasi pantai dan setuju karena pemberhentian proyek ini maka akun tersebut sering disebut aktor lain.

2. Word Cloud

Berikut ini merupakan hasil Word Cloud yang telah diproses pada Rapid Miner dan didapatkan gambar sebagai berikut:

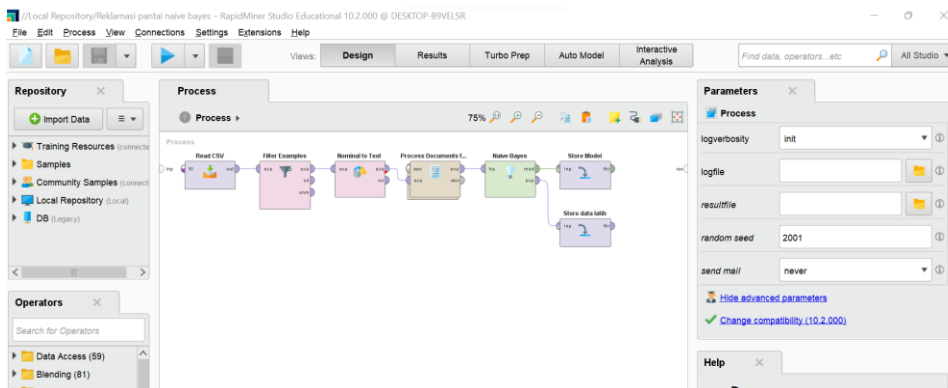
ExampleSet



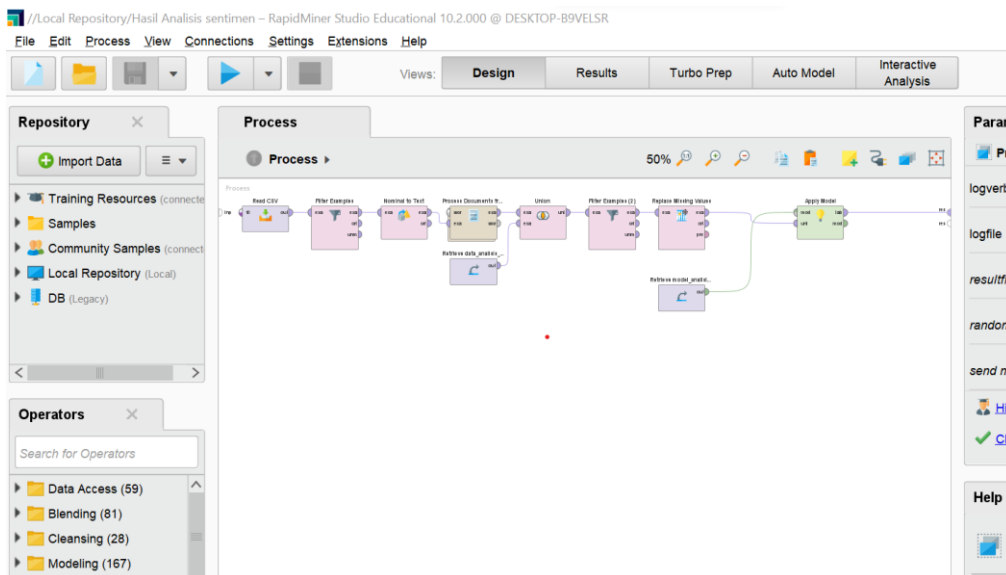
Gambar Visualisasi World Cloud

Pada proses visualisasi Word Cloud diambil 3 kata dari 30 kata teratas. Mengambil 3 kata teratas karena pada 1 dan 2 kata teratas merupakan keyword dari kata tersebut dan hal itu pasti sudah sering disebut oleh netizen, oleh karena itu hanya mengambil dari 3 kata dari 30 kata teratas. Pada gambar di atas yang paling banyak disebut yaitu kata “Jakarta”, “Indonesia”, “Negara”, “Orang”. Kata tersebut memiliki visualisasi yang besar karena sering disebut dalam komentar netizen. Berdasarkan hasil visualisasi Word Cloud di atas dapat ditarik kesimpulan bahwa reklamasi pantai di Jakarta apabila proses pembangunan dilanjutkan kemungkinan dihuni oleh rakyat asing (cina) karena rakyat tersebut kaya sehingga mengakibatkan rakyat miskin atau pribumi tidak dapat menguni hasil dari proyek reklamasi pantai di Jakarta.



3. Sentiment Analysis



Pada gambar di atas merupakan proses analisis sentiment menggunakan algoritma Naïve Bayes. Langkah awal sebelum menggunakan algoritma Naïve Bayes yaitu dengan mengimpor file yang sudah diberi label positif dan negatif sebanyak 300 baris. Langkah selanjutnya seperti pemrosesan sebelumnya kemudian ditraining menggunakan algoritma Naïve Bayes untuk mengetahui prediksi baris yang belum diberi label negatif atau positif. Setelah ditraining menghasilkan 2 data yaitu data model (model_analisis_sentimen) dan data latih (data_analisis_sentimen). Pada data tersebut akan digunakan pada langkah berikutnya.



Gambar di atas merupakan langkah terakhir dari pemrosesan analisis sentiment menggunakan algoritma Naïve Bayes. Pada tahapan pertama masih menggunakan tahapan seperti biasa yaitu mengimpor data yang sudah diberi label sebanyak 300 baris kemudian mengkonverikan data ke text terlebih dahulu. Pada hasil data latih dari algoritma Naïve Bayes akan digabung dengan data yang diberi label 300 baris menggunakan operator union. Setelah digabung akan mengambil data sentiment yang masih belum diberi label dengan menggunakan operator filter example dimana mengambil atribut sentiment dan filter is missing. Jika terdapat data yang kosong maka akan dihapus menggunakan operator missing values. Selanjutnya data yang sudah bersih dan model dari hasil proses algoritma Naïve Bayes akan diapply dengan operator apply model untuk mengetahui prediksi sentimen apakah bernilai positif atau negatif.

Open in  Turbo Prep  Auto Model Filter (345 / 345 examples): all

Row No.	Sentimen	prediction(S...	confidence(...	confidence(...	text
6	?	Positif	0	1	najwa sihab
7	?	Negatif	1	0	reklamasi kalo stop pulau gedung diapain aneh p
8	?	Positif	0	1	kadang bingung kalo teriak korbanya rakyat
9	?	Negatif	1	0	heran gabenar sekarangkalo teliti
10	?	Positif	0	1	pandangan pemikiran digiring pelan pelan
11	?	Positif	0	1	wajah indonesia sesungguhnya dimana kepasti
12	?	Positif	0	1	papa reklamasi kalo tujuannya mencegah amblas
13	?	Negatif	1	0	udah mbak nana bahas bikin runyam ujung ntar s
14	?	Positif	0	1	pulau udah keren stop gitu
15	?	Positif	0	1	bangun rumah keren
16	?	Positif	0	1	menghabiskan uang negara
17	?	Negatif	1	0	liat orang goblok komentar pinter ngerti liat mend
18	?	Positif	0	1	kalo udah ngapain stop pencitraan

Gambar di atas merupakan hasil prediksi sentiment menggunakan algoritma Naïve Bayes yang bernilai positif atau negative. Untuk hasil yang bernilai positif seperti pada kalimat “bangun rumah keren” pada baris ke-14, sedangkan kalimat bernilai negatif yaitu pada kalimat “reklamasi kalo stop pulau gedung diapain aneh pemerintah jaman” pada baris ke-7. Namun, pada hasil prediksi tersebut data yang dihasilkan tidak semuanya akurat, terdapat text yang seharusnya bernilai negatif namun diprediksi bernilai positif seperti pada baris ke-16 dengan kalimat “Menghabiskan uang Negara”. Kalimat tersebut merupakan kalimat yang mengandung nilai negatif.

E. Kesimpulan

Berdasarkan hasil Analisis Sentimen komentar Youtube terhadap Pemberhentian Reklamasi Pantai Menggunakan Algoritma Naïve Bayes yaitu proyek pembangunan reklamasi pantai di Teluk Jakarta menghasilkan nilai negatif. Hasil dari nilai negatif ini yaitu bahwa proyek pembangunan reklamasi pantai dihentikan karena memiliki dampak buruk untuk kedepannya apabila proyek reklamasi pantai ini dilanjut. Dampak yang terjadi berdasarkan analisis Word Cloud yaitu jika pembangunan ini selesai, kemungkinan yang



**FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO SEMARANG**

Jl. IMAM BONJOL NO. 207 SEMARANG TELP. 024-3575915, 024-3575916

menghuni adalah orang asing dari cina karena orang asing tersebut kaya sedangkan pribumi tidak bisa menghuni kawasan tersebut karena harganya dipastikan sangat mahal.



**FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO SEMARANG**

JL. IMAM BONJOL NO. 207 SEMARANG TELP. 024-3575915, 024-3575916

Ketentuan:

1. Judul dituliskan dengan jumlah maksimal 12 kata
2. Bagian pendahuluan terdiri atas tiga paragraf, meliputi:
 - a. Paragraf 1 merupakan penjelasan tentang **fenomona / fakta / latar belakang** yang berkaitan dengan topik yang telah ditentukan
 - b. Paragraf 2 menjelaskan tentang **masalah dan akibat** / efek yang muncul karena masalah
 - c. Paragraf 3 menjelaskan tentang **tujuan** dilakukannya analisis word cloud dan sentiment analysis
3. Dataset hasil dari Netlytic, dataset hasil processing dan dataset labelling, file RapidMiner di upload di Google Drive
4. Algoritma yang digunakan untuk sentiment analysis adalah Naive Bayes

Catatan:

1. File dikumpulkan dalam format **.pdf** dan disubmit melalui **Kulino**
2. Penamaan file adalah **UTS_AMS_NIMxxxxxx**. Contoh: **UTS_AMS_02237**
3. Tidak perlu ditambahkan cover halaman atau halaman yang sejenis
4. Dikumpulkan di Kulino paling lambat **Jumat, 3 November 2023 pukul 12.00 WIB**
5. *Plagiarism will not be tolerated*